

Lowering the learning curve for declarative programming: a Python API for the IDP system

Joost Vennekens

Dept. Computer Science, KU Leuven
Technology Campus De Nayer, Sint-Katelijne-Waver, Belgium

Abstract. Programmers may be hesitant to use declarative systems, because of the associated learning curve. In this paper, we present an API that integrates the IDP Knowledge Base system into the Python programming language. IDP is a state-of-the-art logical system, which uses SAT, SMT, Logic Programming and Answer Set Programming technology. Python is currently one of the most widely used (teaching) languages for programming. The first goal of our API is to allow a Python programmer to use the declarative power of IDP, without needing to learn any new syntax or semantics. The second goal is allow IDP to be added to/removed from an existing code base with minimal changes.

1 Introduction

While declarative systems may not be inherently more difficult to learn than, e.g., an object-oriented programming language, they are still often perceived as such, because many programmers are unfamiliar with their syntax, semantics or paradigm. This leads in turn to a chicken-and-egg problem, where programmers do not learn this technology, because companies are reluctant to adopt it, because there are not enough programmers who know the technology readily available.

In this paper, we propose to tackle both problems by means of an API that allows a declarative knowledge base (KB) to be used from within a well-known imperative host language. Our first goal is to integrate KB functionality into the host language as seamlessly as possible. In this way, it should be possible to use the knowledge base to prototype a single component of a large system, without affecting the rest of the code base. The second goal is to have a very low learning curve for the API. We achieve this by using as much as possible the syntax of the host language, and by requiring no more background in declarative languages than an introductory course on classical logic. We therefore expect our API to be immediately usable by, e.g., bachelor students in a typical CS curriculum.

In more detail, we postulate these guidelines for our development of the API:

- The interaction between the KB and host language should be done through standard objects of the host language.
- The need to learn KB-specific terminology should be kept to a minimum.
- It should be as easy as possible to replace the KB by a piece of host language code (or, vice versa, to replace a piece of host language code by a KB).

A typical use for our API will be to off-load specific computational problems (e.g., detect connected components in a graph, find a permissible allocation of resources to jobs) to the KB, thereby avoiding the need to implement a specific algorithm and thus arriving at a working prototype more quickly. In such an early prototype, the modular, declarative nature of the KB will be particularly useful, because of its ability to easily cope with additional changes to the specification. Once the program has reached a certain level of maturity, it can of course be profiled to see whether all of the KB components meet the performance requirements. Whenever this is not the case, the KB can be replaced by a dedicated algorithm with minimal impact on the rest of the code.

As a host language, we use Python (in particular, version 2.7). Given our stated goals, this is the most obvious choice: “[a]t the time of writing (July 2014), Python is currently the most popular language for teaching introductory computer science courses at top-ranked U.S. departments.”¹ We assume familiarity with the basics of Python throughout this paper. The KB system that we use will be discussed in Section 2. Section 3 then discusses our interface between host language and KB, which we validate by means of some examples in Section 4. In Section 5, we give some brief notes on the implementation of our API. Finally, Section 6 discusses some related work, in particular, other approaches that integrate a declarative knowledge base into an imperative language.

2 KB system

As an underlying KB system, we will use IDP (*Imperative Declarative Programming*)² [2], which combines techniques from SAT solving, Logic Programming and Answer Set Programming (ASP). It has performed well in previous ASP competitions, e.g., narrowly finishing second after Clasp in the System Track of 2011³. IDP has a number of properties that fit well with our goals of achieving both a tight integration with the host language and a low learning curve.

Input language. IDP uses a language that is a conservative extension of classical first-order logic (FO). Because most students of computer science are familiar with FO, this means that the learning curve for a large part of IDP’s input language consists only of learning a particular ASCII syntax for FO. One of the ways in which IDP extends FO is by adding *inductive definitions* [3]. Because such definitions cannot, in general, be expressed in FO, this is a real extension of the language. Moreover, since inductive datatypes (lists, trees, ...) are very common in computer programs, this feature will prove useful in our API.

Inference. In addition to its input language, a second useful property of IDP is that it aims to support a variety of different inference tasks. Particularly useful in the context of this article is the task of *finite model expansion*. As pointed out in [6], modal expansion for FO captures the complexity class NP, thereby

¹ <http://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-us-universities/fulltext>

² <https://dtai.cs.kuleuven.be/software/idp>

³ <https://www.mat.unical.it/aspcomp2011/>

covering the kind of tasks that we would like to off-load to a declarative KB. Moreover, [7] have further demonstrated that model expansion is a key task when using declarative methods to build modular software systems.

2.1 FO: syntax and semantics

We briefly recall the standard syntax and semantics of FO. A *vocabulary* Σ consists of a set of function symbols, each with an associated arity n , and a set of predicate symbols, also each with an arity n . A function with arity 0 is called a *constant*. A *term* is either a constant, a variable or an expression $f(t_1, \dots, t_n)$ where f is an n -ary function symbol and the t_i are terms. An *atom* is an expression $P(t_1, \dots, t_n)$, with P an n -ary predicate and the t_i again terms. A *formula* is either an atom or an expression $\psi \vee \phi$, $\psi \wedge \phi$, $\neg \phi$, $\forall x : \phi$, or $\exists x : \phi$, where ψ, ϕ are formulas and x is a variable. As usual, $\phi \Rightarrow \psi$ abbreviates $\neg \phi \vee \psi$ and $\phi \Leftrightarrow \psi$ stands for $(\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi)$. A *sentence* is a formula without free variables and a *theory* is a finite set of sentences.

The standard semantics of FO is defined in terms of *structures* for a vocabulary Σ . Each structure S consists of a domain D and a mapping of:

- Each n -ary predicate symbol P in Σ to an n -ary relation $R \subseteq D^n$
- Each n -ary function symbol f in Σ to an n -ary function $f : D^n \rightarrow D$

The *satisfaction* relation \models is defined between structures for a vocabulary and theories of this vocabulary (or a subvocabulary thereof) by the usual induction. When $S \models T$, we also say that the structure S is a *model* of T .

2.2 FO in the IDP system

IDP uses the standard concepts of vocabularies, theories and structures. Each of these has a specific syntactic representation. For example, a map coloring problem can be described in the following vocabulary V .

```
vocabulary V {
  type Color
  type Area
  Border(Area,Area)
  Coloring(Area): Color
}
```

FO	IDP	Python
\wedge	<code>&</code>	<code>and</code>
\vee	<code> </code>	<code>or</code>
\neg	<code>~</code>	<code>not</code>
\Rightarrow	<code>=></code>	<code>not present</code>
$=$	<code>=</code>	<code>==</code>
\neq	<code>~=</code>	<code>!=</code>
\forall	<code>!</code>	<code>all</code>
\exists	<code>?</code>	<code>any</code>

Fig. 1. ASCII syntax of IDP.

As can be seen here, IDP in fact uses a typed variant of first-order logic. The first two statements define two types (which can be seen as unary predicates), whereas the last two statements define, respectively, a predicate and a function.

The following theory T in vocabulary V consists of a single sentence, expressing that neighboring areas must have a different color. Fig. 1 shows the ASCII symbols that are used in IDP to represent the logical connectives.

```

theory T : V {
  !a b: Border(a,b) => Coloring(a) ~= Coloring(b).
}

```

IDP is usually able to automatically derive the types of variables, based on the type declarations in the definition of the vocabulary. This information is important because most of its inference tasks require IDP to first *ground* (part of) the theory. Instead of depending on IDP's automatic type derivation, it is also possible to explicitly declare the type of a variable, e.g.:

```

!a [Area] b [Area]: Border(a,b) => Coloring(a) ~= Coloring(b).

```

A structure S for a vocabulary V is represented by an enumeration (1) of the values that belong to each type, (2) of the tuples that belong to each predicate, and (3) of the mapping of tuples to values that is made by each function.

```

structure S : V {
  Area  = { Belgium; Holland; Germany; }
  Color = { Blue; Red; Green; }
  Border = {(Belgium,Holland); (Belgium,Germany); (Holland,Germany)}
}

```

This structure S for the vocabulary V interprets only part of the vocabulary V . In particular, the function *Coloring* is not interpreted. One of the inference tasks supported by the IDP system is that of *model expansion*: given a structure S' for a subvocabulary Σ' of the vocabulary Σ of a theory T , compute a structure S for the remaining symbols $\Sigma \setminus \Sigma'$ such that $(S' \cup S) \models T$.

IDP exposes its functionality by means of an API in the Lua scripting language. The command `printmodels(modelexpand(T,S))` performs the model expansion task for the above structure S and theory T , resulting in the output:

```

Coloring = {"Belgium"->"Red";"Germany"->"Blue";"Holland"->"Green"}

```

By default, a single model expansion is computed, but it is also possible to compute several or all of them. A special case of model expansion occurs when the initial structure S already interprets the entire vocabulary V of the theory T . In this case, it reduces to checking whether $S \models T$.

3 Interfacing with the KB System

This section presents our API for using the IDP KB system from within Python.

3.1 Vocabularies and structures

All interaction with the IDP system is done through objects of the IDP class. Each such object represents a knowledge base consisting of a triple (V, S, T) of a vocabulary V , a structure S and a theory T . It can be created as follows:

```
kb = IDP()
```

The following methods add symbols to the vocabulary of the KB:

```
kb.Type(name [, interpretation])
kb.Constant(typed_name [, interpretation])
kb.Function(typed_name [, interpretation])
kb.Predicate(typed_name [, interpretation])
```

As in IDP itself, the `typed_name` of a predicate is a string of the form `Foo(Type1, ..., Type2)` and that of a function is `Foo(Type1, ..., Type2): Return_type`. Because constants are identical to 0-ary functions, their `typed_name` has the form `Foo : Type`. Once a symbol `Foo` has been added to the vocabulary of a knowledge base `kb`, it can thereafter be referred to as `kb.Foo`.

In addition to declaring a function/predicate symbol σ (i.e., adding it to the vocabulary of the KB), it is also possible to immediately extend the structure S with a particular interpretation σ^S for σ . This is done by adding this interpretation as a second argument. The interpretation of a type must be a set (or list) of values; that of a constant must be a single value; that of a function with arity ≥ 1 must be mapping (e.g., a dictionary); and that of a predicate must be a set/list of tuples of the correct arity (for predicates with arity 1, a set of simple values is also allowed). Obviously, the typing of the symbols must be respected.

Instead of initialising the interpretation of a symbol upon construction, it is also possible to first declare a symbol and then later use the assignment operator to provide an interpretation for it. We illustrate using (part of) the graph coloring example of Sec. 2.2.

```
color = IDP()
color.Type("Color", ["Blue", "Red", "Green"])
color.Type("Area", ["Belgium", "Holland", "Germany"])
color.Predicate("Border(Area, Area)")
color.Border=[("Belgium", "Holland"), ("Belgium", "Germany"),
              ("Holland", "Germany")]
color.Function("Coloring(Area): Color")
```

The “logical” objects that are thus created implement a number of common Python interfaces, allowing them to act as Python programmers would expect. A *relation* is, in mathematical terms, a set of tuples. Its natural counterpart is a Python *MutableSet* object⁴ (i.e., a set which allows adding/removing of elements). The following interactive session demonstrates some standard usages.

```
>>> "Belgium" in color.Area
True
>>> color.Area.add("Austria")
>>> for x in color.Area:
```

⁴ <https://docs.python.org/2/library/collections.html#collections.Set>

```
...     print x,
...
Holland Austria Germany Belgium
```

In addition to the standard *MutableSet* functionality, relations are also *callable*⁵, so that we may also use the standard FO notation for checking membership.

```
>>> color.Area("Belgium")
True
```

A *function* is a *Mapping*⁶, that is, an object that maps each tuple of values in its domain to a value in its range. Some standard usages are:

```
>>> color.Coloring.keys()
['Austria', 'Germany', 'Holland', 'Belgium']
>>> color.Coloring["Belgium"]
'Blue'
```

As with predicates, function objects are *callable* to allow for the more FO-like:

```
>>> color.Coloring("Belgium")
'Blue'
```

3.2 Formulas and definitions

In keeping with our goal of achieving a low learning curve, formulas are written in Python syntax. An overview is shown in Fig. 1. The Python language has the standard boolean operators **and**, **or** and **not**. In addition, it also has the functions **all** and **any**, which may be applied to lists of boolean values to return the conjunction/disjunction of these values. The latter two functions, together with Python's list comprehension syntax, can be used as universal/existential quantification. The list comprehension syntax also has an optional **if** part, which may be used to represent the common pattern of a universally quantified implication:

$$\begin{array}{ll}
 \forall x[Type] : \phi & \text{all}(\phi \text{ for } x \text{ in Type}) \\
 \exists x[Type] : \phi & \text{any}(\phi \text{ for } x \text{ in Type}) \\
 \forall x[Type] : \phi \Rightarrow \psi & \text{all}(\psi \text{ for } x \text{ in Type if } \phi)
 \end{array}$$

In the graph coloring problem, we need to express the following property:

$$\forall a b : \text{Border}(a, b) \Rightarrow \text{Coloring}(a) \neq \text{Coloring}(b) \quad (1)$$

Section 2.2 already presented the IDP syntax for this. In Python, we can write the same property as:

⁵ <https://docs.python.org/2/library/functions.html#callable>

⁶ <https://docs.python.org/2/library/collections.html#collections.Mapping>

```
all(color.Coloring(a) != color.Coloring(b)
    for (a,b) in color.Border)
```

Note that this is just a normal Python expression, which we can, e.g., just type into the interactive terminal. This expression evaluates to `True` precisely when property (1) is satisfied. We can make the KB aware of such a constraint by means of its `Constraint` method, which takes a string as its argument. So, the following code adds the above constraint to our graph coloring KB:

```
color.Constraint("all(Coloring(a) != Coloring(b)
                    for (a,b) in Border)")
```

The string argument is completely identical to the Python expression we saw above, with one exception: the predicates/functions are simply called `Coloring`, `Area` and `Border`, instead of `color.Coloring`, `color.Area` and `color.Border`. This is because, just as a theory can only contain symbols that appear in its vocabulary, the constraints that are added to a KB always refer to the symbols of that KB.

An alternative formula, which is equivalent to the one above, quantifies over $Area \times Area$ and uses an `if`-expression to check for membership in `Border`:

```
color.Constraint("all(Coloring(a) != Coloring(b)
                    for a in Area for b in Area if (a,b) in Border)")
```

3.3 Functional interfacing

In keeping with our goal of making the API easy to use, the programmer does not need to explicitly invoke the IDP system. This avoids the need to learn new functions or new terminology, and reduces the possibility of bugs. Instead, invocation of the IDP system happens “automatically” in the following circumstances:

- Symbols that have been declared, but for which no interpretation has been provided, are automatically assigned a valid interpretation (in accordance with FO semantics) when their content is inspected. In other words, IDP is used as an *oracle* to lazily fill in the interpretation of any declared symbols for which the user does not provide one herself. This is done in a way that the interpretations of all symbols together constitutes an FO model of the constraints, i.e., a model expansion task is performed. If the constraints admit multiple models, one is chosen arbitrarily.
- The KB object has an attribute `satisfiable`, which is automatically set to `True/False` (depending on whether the KB is satisfiable) when the user converts it to a boolean (either explicitly with `bool(.)` or by use in an `if`-statement).

In the previous section, we declared a function `color.Coloring` without adding an interpretation for this function. Therefore, the IDP system will be invoked to compute a coloring of our graph if we execute the following code:

```

for x in color.Area:
    print "Area %s has color %s" % (x, color.Coloring[x])

```

Note that if we *had* added an interpretation for the function `Coloring` before executing this code, then this `for`-loop would still continue work as expected. This is an important property, because it allows us to change whether a relation/function is computed by the KB base or by native Python code, without having to adjust the code that makes use of it. In the latter case, the call `color.Coloring[x]` will just retrieve the pre-computed coloring stored within the `color` object, without invoking IDP.

Similarly, if a coloring is not provided, then the following code will test whether a given graph *can be* colored, whereas if it is provided, the same code checks whether it is indeed valid.

```

if color.satisfiable:
    print "The graph can be colored"

```

3.4 Inductive definitions

An important feature of the IDP system is its ability to handle *inductive definitions*. It uses a rule-based syntax for representing such definitions, in which, e.g., the transitive closure TC of a graph G can be defined as follows

$$\left\{ \begin{array}{l} \forall x, y : TC(x, y) \leftarrow G(x, y). \\ \forall x, y : TC(x, y) \leftarrow \exists z : G(x, z) \wedge TC(z, y). \end{array} \right\}$$

Note that the arrow symbol here is not material implication, but a special symbol that denotes a “case” in an inductive definition. Such an inductive definition is interpreted under the well-founded semantics [9], which in the case of a positive definition (such as the one above) boils down to a least-fixpoint construction. Each rule of such a definition represents a single case in which the defined predicate holds. In our Python API, we use a lambda-expression to represent such a case.

```

kb.Define([( "TC(Node,Node)", "lambda x,y: G(x,y)" ),
            ( "TC(Node,Node)", "lambda x,y:
                                any(G(x,z) and TC(z,y) for z in Node)" )])

```

This both declares the predicate `TC` and defines it in terms of the “parameter” `G`. For definitions consisting of a single rule, a simpler syntax is also allowed:

```

kb.Define("TC(Node,Node)",
          "lambda x,y: G(x,y) or
          any(G(x,z) and TC(z,y) for z in Node)")

```


Similar to how an argument of `kb.Constraint(.)` can also be used as a simple boolean Python expression, the above lambda-expression can be used to compute the transitive closure of G by an explicit least-fixpoint computation:

```
def lfp(f, x=[]):
    y = f(x)
    return y if y == s else f(x)

node_pairs = [(x,y) for x in kb.Node for y in kb.Node]
TC = lfp(lambda T: filter(lambda x,y: kb.G(x,y) or
    any(kb.G(x,z) and T(z,y) for z in Node), node_pairs))
```

An advantage of using IDP is that the definition of TC can then not only be used to compute the transitive closure of a given graph, but also to, e.g., compute a graph that would have a given relation as its transitive closure. In addition, IDP not only supports positive inductive definitions, but also non-monotone inductive definitions (such as the standard definition of the relation “ \models ” in FO), for which a simple least-fixpoint construction does not work. Non-recursive definitions (which are equivalent to a standard FO equivalence) are also allowed in IDP. In the latter case, we can of course choose whether to use the `Define` or `Predicate` method of our API.

IDP can be configured to use XSB Prolog⁷ to speed up certain computation with definitions. We always use this option in the experiments below.

4 Experiments

This section presents two examples of our API, with a particular focus on demonstrating that the integration into the surrounding Python code can be done in a natural way.

4.1 Sudoku

The first example is a Sudoku solver. A Sudoku grid consists of $9 \times 9 = 81$ cells.

```
sud = KB()
sud.Type("Cell", range(81))
```

The grid is divided into in rows, columns and nine small 3×3 squares.

```
sud.Predicate("SameRow(Cell, Cell)",
    [(i, j) for i in sud.Cell for j in sud.Cell
     if i / 9 == j / 9 ])
sud.Predicate("SameCol(Cell, Cell)",
    [(i, j) for i in sud.Cell for j in sud.Cell
     if i % 9 == j % 9 ])
```

⁷ <http://xsb.sourceforge.net/>

```

sud.Predicate("SameSmallSq(Cell, Cell)",
[ (i, j) for i in sud.Cell for j in sud.Cell
  if (i%9)/3 == (j%9)/3 and (i/9)/3 == (j/9)/3 ])

```

Here, the Python list comprehensions compute an enumeration of these relations, by iterating over all tuples in the Cartesian product of the argument types and checking a certain condition for each tuple. Alternatively, we can make IDP do this work, by Defining the predicates with the appropriate *lambda*-expressions.

```

sud.Define("SameRow(Cell, Cell)",
          "lambda i, j: i / 9 == j / 9")

```

The cells must be filled with integers from 1 to 9. It will be convenient to represent an empty cell by the number 0, leading to the following type *Number*.

```

sud.Type("Number", range(10))

```

We make use of two functions that map cells to numbers: one records the problem statement and the other its solution. The problem statement comes from a list of numbers, that we convert to a dictionary by means of `zip` operation.

```

s=[ 8,5,0, 0,0,2, 4,0,0,
    7,2,0, 0,0,0, 0,0,9,
    0,0,4, 0,0,0, 0,0,0,
    ... .. ]
sud.Function("Given(Cell):Number",
            dict(zip(range(len(s)), s)))
sud.Function("Sol(Cell): Number")

```

The function `Sol` will be computed by IDP, in accordance with the rules of sudoku. First, we state the difference constraint on the appropriate cells.

```

sud.Define("Diff(Cell,Cell)", "lambda x,y: x != y and
    (SameRow(x,y) or SameCol(x,y) or SameSmallSq(x,y))")
sud.Constraint("all(Sol(x) != Sol(y) for (x,y) in Diff)")

```

Next, we state that the solution must match the problem statement on all non-empty cells (i.e., those $\neq 0$), and that it should fill in all cells.

```

sud.Constraint("all(Sol(x) == Given(x)
    for x in Cell if Given(x) != 0)")
sud.Constraint("all(Sol(x) != 0 for x in Cell)")

```

With this, the sudoku problem is completely specified. The following code passes the `sud.Sol` object to a function that pretty-prints the sudoku. Only at the start of the `for`-loop in this function is the solution actually computed.

```

def show(grid):
    row = -1
    for x in grid.keys():

```

```

sys.stdout.write(' ' if x % 3 == 0 else '')
if x % 9 == 0:
    row += 1
    print ('\\n' if row % 3 else '')
print (str(grid[x]) + " "),

show(sud.Sol)

```

We remark that, because we use valid Python expressions to assert constraints, we can use the same expressions to check that the output indeed satisfies the constraints. For instance, at the interactive Python terminal:

```

>>> all(sud.Sol(x) != sud.Sol(y) for (x,y) in sud.Diff)
True

```

Peter Norvig has published a sudoku solver written entirely in Python using constraint solving techniques⁸. Not counting whitespace and comments, his code to compute solutions is about 40 lines, whereas the code we have presented in this section is 12 lines. Moreover, it is easy to replace the existing `solve` function in his code by a call to our knowledge base. This requires two small transformations: first, in Norvig's code, input grids are given in the format of a single string, where an empty cell is represented by a dot; second, he produces output in the form of a dictionary in which the keys are strings of the form Xn with $X \in A..I$ representing the row and $n \in 1..9$ the column.

```

def solve(sud, grid):
    """ Solve sudoku "grid" using the KB "sud". """
    def translate(n):
        rows      = 'ABCDEFGHI'
        digits     = '123456789'
        return rows[n/9] + digits[n%9]
    sud.Given = dict(zip(range(81),
        [int(x) for x in grid.replace('.', '0')]))
    return dict(zip(map(translate, sud.Sol.keys()),
        map(str, sud.Sol.values()))))

```

When it comes to runtime, our version is significantly slower than the original on Norvig's test set, averaging 0.27s per sudoku versus 0.008s.

At the other end of the spectrum, we can also compare to a naive generate-and-test approach. Using Python's powerful `itertools` library, this can also be implemented in about 10 lines of code. (The code used to test whether a solution is correct can of course use the same syntactical expressions as those which we passed to our API.) However, the runtime of such a program is very poor: for a sudoku with just 5 empty cells (for comparison, a typical sudoku has around 50 to 60), it already takes over a minute to find a solution.

Our main conclusions from this experiment are that, at least in this case:

⁸ <http://norvig.com/sudoku.html>

- Our API can handle, with a limited amount of overhead, the input/output format that a typical Python programmer would use;
- Our API can be used to develop useful functionality in significantly fewer lines of code (12 versus 40) than a clever Python implementation. In fact, it takes only as many lines of code as a naive generate-and-test algorithm.
- Even though our API is significantly slower than a clever Python algorithm, it still vastly outperforms the naive generate-and-test approach.

4.2 Working with graphs

The following class `GraphKB` extends the generic IDP Knowledge Base class with some specific functionality for working with undirected graphs. When constructing such a `GraphKB`, the nodes of the graph can be initialised by means of a given set and the edges by means of an adjacency list. The predicate `Edge` is Defined as the symmetric closure of the adjacency list. This class also offers a convenience method to define the transitive closure of relations over this graph.

```
class GraphKB(IDP):

    def __init__(self, nodes=[0], adj_list=[]):
        super(GraphKB, self).__init__()
        self.Type("Node", nodes)
        self.Predicate("Adjacent(Node,Node)", adj_list)
        self.Define("Edge(Node,Node)",
                    "lambda x,y: Adjacent(x,y) or Adjacent(y,x)")

    def add_TC(self, original, tc_name):
        formula = "lambda x,y: {0}(x,y) or any({0}(x,z)
            and {0}(z,y) for z in Node)".format(original)
        self.Define(tc_name + "(Node, Node)", formula)
```

We can now check if a given adjacency list describes a fully connected graph:

```
connected = GraphKB(nodes, adj)
connected.define_TC("Path", "Edge")
connected.Constraint("all(Path(x,y)
                    for x in Node for y in Node)")
if connected.satisfiable:
    print "Graph is fully connected"
```

We can use a similar KB to count the number of connected components in the graph. We do this by selecting a single representative from each component (its “Root”) and then counting the number of these representatives.

```
comp = GraphKB(nodes, adj)
comp.define_TC("Path", "Edge")
comp.Predicate("Root(Node)")
```

```

comp.Constraint("all(any(Path(r,x) for r in Root)
                    for x in Node if not Root(x))")
comp.Constraint("not any(Path(x,y)
                        for x in Root for y in Root if x != y)")
print "Number of components: {0}".format(len(comp.Root))

```

For a graph with 1000 nodes in 86 connected components, this program takes 23s to count the components. By comparison, the popular NetworkX Python library⁹ is two orders of magnitude faster, taking only 0.2s.

In graph theory, an undirected graph is called a *tree* if it is connected and does not contain cycles. When checking for a cycle in an undirected graph, we of course have to exclude the trivial two-node cycles that would result from traversing the same undirected edge in both directions. This in fact makes it easier to use IDP to check that there *is* a cycle, than to check that there *is not* one. The following knowledge base tries to guess the direction in which to traverse each edge in order to produce a cycle. If it is unsatisfiable, there are no cycles.

```

cyclic = GraphKB()
cyclic.Predicate("Traverse(Node,Node)")
cyclic.Constraint("all(Edge(x,y) for (x,y) in Traverse)")
cyclic.Constraint("not any(Traverse(y,x)
                        for (x,y) in Traverse)")
cyclic.define_TC("TravTC", "Traverse")
cyclic.Constraint("any(TravTC(x,x) for x in Node)")

```

We can now combine the two knowledge bases to check whether a given adjacency list indeed describes a tree.

```

def is_tree(adj_list):
    cyclic.Adjacent = adj_list
    connected.Adjacent = adj_list
    return (bool(connected.satisfiable)
            and not bool(cyclic.satisfiable))

```

This example illustrates how additional functionality can be built on top of the KB objects of our API. In addition, the ability to combine the results of calls to different KBs also allows us to implement functionality that would be harder to implement in a single IDP KB.

5 Implementation

The implementation of our API and the examples are available for download.¹⁰ Interfacing with the IDP system is currently done in a decoupled way: when the

⁹ <https://networkx.github.io/>

¹⁰ <https://bitbucket.org/joostv/pyidp/admin>

API detects that the IDP system needs to be called, it prepares a text file with the appropriate vocabulary, structure and theory, then calls the IDP system as an external process and parses its output. The results of this call are cached, so that the IDP system will not be invoked again until the KB changes.

6 Related work

There is already a long history of work attempting to close the gap between imperative and declarative programming [1]. We briefly compare our approach to some recent work in this area.

In [8], an approach is presented in which a constraint solver is not added to a single host language, but can be used in the development of a domain-specific language in Racket. Like ours, the motivation behind this work is to allow the power of declarative systems to be more widely used. However, their approach differs, because they count on an intermediary—the designer of the domain-specific language—to hide the complexity of the declarative system, whereas our approach focuses on creating an interface that is natural enough to offer KB functionality directly.

In [4], a constraint solver is integrated into the Scala language. As ours does, their approach reuses the syntax of the host language to interface with the declarative system. A key difference is that, in their approach, the programmer is explicitly manipulating, combining and solving constraints, which makes the constraint solver more present in the eventual source code. A second difference is of course that Scala currently appears to be less widely known than Python.

In [5], a reasoner for FO extended with transitive closure is integrated into Java. Their KB language is therefore very similar to (but more restricted than) that of IDP. When it comes to the integration in Java, there are two main differences to our approach. First, the declarative knowledge is not written in expressions in the host language, but in a separate language (the Alloy-like JFSL [10]). Second, the integration into Java is done in an object-oriented way: the programmer defines classes in which formulas are added as, among others, class invariants, method pre-/postconditions and frame conditions. In comparison, our Python API seems more lightweight, since it does not require an object-oriented approach. When it comes to computational performance, [5] reports good results, which our implementation is not able to match.

In summary, our approach fills the niche of an easy-to-learn quick prototyping API, that, due to Python’s current popularity, may speak to a large audience.

7 Conclusions and future work

When prototyping an application, a programmer may encounter a computational subproblem for which it would be cumbersome to develop a specific algorithm. The aim of our API is to allow such gaps to be declaratively stopped with as little effort as possible. As we have seen, our API might allow a feasible solution to be produced in only as many lines of code as an (infeasible) naive

generate-and-test algorithm. Our use of standard Python objects such as sets and mappings means that no elaborate setup code is required to plug the KB into an existing code base, while our use of standard Python expressions for constraints and definitions leads to a low learning curve. In addition, both these properties also make it easier to eventually remove the KB if a more efficient solution is required: the same KB that first generated the solution, can later be used to check its correctness, or its constraints may simply be recuperated in the form of Python `assert`-statements.

To prevent changing/removing the KB from leading to code changes elsewhere, our API makes all calls to the IDP system automatically, whenever they are needed. This has the additional benefit of simplifying the API and not forcing the programmer to learn new terminology. A downside is that it is harder for the programmer to keep track of what is happening when in the program.

Our current implementation of the API is naive in its interfacing with the IDP system, which happens by passing text files (built each time from scratch) to an external process. A better integration, which exploits the Lua interface of IDP, might offer a significant reduction in runtimes. However, since we mainly intend our API to be used in prototyping, this issue might not be pressing. Another consequence, which may be more severe, is that programs written in our API are currently hard to debug: it may be necessary to manually inspect the text file that was passed to the IDP system (in debug-mode, the API always sends this to standard output). However, this requires the user to be at least somewhat familiar with IDP input syntax, which is something we aimed to avoid.

Our validation of the API currently consists only of examples that we have implemented ourselves. A better test would involve Python programmers who have no knowledge of IDP or indeed any declarative system. However, better debugging facilities seem necessary for such a trial to be successful.

References

1. Krzysztof R. Apt and Andrea Schaerf. Programming in alma-0, or imperative and declarative programming reconciled. In *FroCos*, 1998.
2. M. Bruynooghe, H. Blockeel, B. Bogaerts, B. De Cat, S. De Pooter, J. Jansen, A. Labarre, J. Ramon, M. Denecker, and S. Verwer. Predicate logic as a modeling language: Modeling and solving some machine learning and data mining problems with IDP3. *Theory and Practice of Logic Programming*, 2014. Accepted.
3. Marc Denecker and Eugenia Ternovska. A logic of nonmonotone inductive definitions. *ACM Transactions on Computational Logic*, 9(2), March 2008.
4. A. Köksal, V. Kuncak, and P. Suter. Constraints as control. In *POPL'12*, 2012.
5. A. Milicevic, D. Rayside, K. Yessenov, and D. Jackson. Unifying execution of imperative and declarative code. In *Proc. 33rd Int'l Conference on Software Engineering (ICSE)*, 2011.
6. David G. Mitchell and Eugenia Ternovska. A framework for representing and solving NP search problems. In *AAAI*, pages 430–435, 2005.
7. Shahab Tasharrofi and Eugenia Ternovska. A semantic account for modularity in multi-language modelling of search problems. In *Proc. FroCos*, 2011.
8. E. Torlak and R. Bodik. Growing solver aided languages with ROSETTA. In *Proc. ACM international symposium on New ideas, new paradigms, and reflections on programming & software*, 2013.
9. A. Van Gelder, K.A. Ross, and J.S. Schlipf. The well-founded semantics for general logic programs. *Journal of the ACM*, 38(3):620–650, 1991.
10. K. Yessenov. A lightweight specification language for bounded program verification. Master's thesis, MIT, 2009.